

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-81527

(43)公開日 平成9年(1997)3月28日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/16	3 7 0		G 0 6 F 15/16	3 7 0 N
3/06	5 4 0		3/06	5 4 0
13/00	3 5 5		13/00	3 5 5

審査請求 未請求 請求項の数15 O L (全 17 頁)

(21)出願番号 特願平8-186269

(22)出願日 平成8年(1996)7月16日

(31)優先権主張番号 08/531239

(32)優先日 1995年9月19日

(33)優先権主張国 米国 (US)

(71)出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州アーモンク (番地なし)

(72)発明者 イアン・ディビット・ジャド

イギリス、エス021 2イー・ジイ、ウィンチェスター、オッターバーン、コールス・メード 33

(74)代理人 弁理士 合田 潔 (外2名)

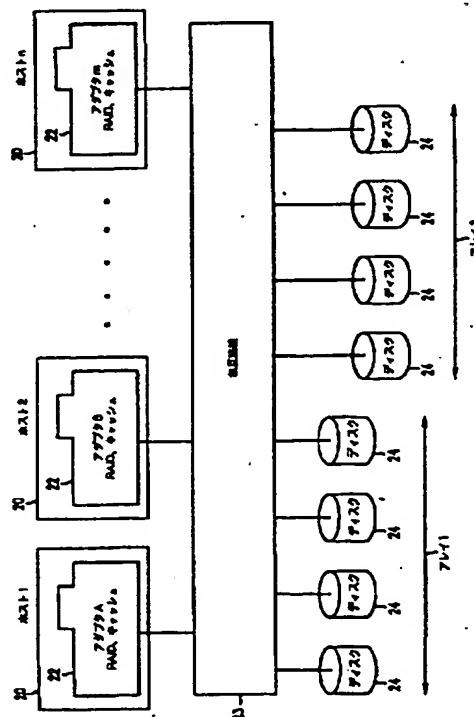
最終頁に続く

(54)【発明の名称】 複数のホスト・コンピュータ・システムにより複数の記憶装置アレイを共有するシステム及び方法

(57)【要約】

【課題】 ホスト及び記憶装置アレイのより大規模な接続性を可能にする、より安価でよりスケーラブルな解決策を提供する。

【解決手段】 各記憶装置アレイ上のデータが、任意のホスト・コンピュータによりアクセスされ得るように、複数のホスト・コンピュータのデータを複数の記憶装置アレイ上に記憶するシステムが提供される。複数のアダプタ・カードが使用され、各アダプタは、指定された記憶装置アレイに対する制御装置機能を有する。システム内の全てのアダプタ間に、アダプタ通信インタフェース (相互接続) が存在する。



【特許請求の範囲】

【請求項1】複数のホスト・コンピュータのデータを複数の記憶装置アレイ上に記憶し、任意の記憶装置上のデータが任意のホスト・コンピュータによりアクセスされ得るようにするシステムであって、

各々がホスト・コンピュータに関連付けられ、指定されたアレイの1次制御を有する、複数のアダプタと、前記アダプタ間のピア・ツー・ピア通信のためのアダプタ通信相互接続と、

を含み、

ホスト・コンピュータから関連アダプタへの、前記アダプタにより本来制御されないアレイに対するデータ・アクセス要求が、前記アダプタ通信相互接続を介して、前記アレイの1次制御を有する前記アダプタに伝達される、システム。

【請求項2】前記アダプタにより本来制御される前記記憶装置を示す識別子を、当該各アダプタ内に記憶する、請求項1記載のシステム。

【請求項3】前記記憶装置アレイの1次制御を有するアダプタを識別する識別子を、当該各記憶装置アレイに記憶する、請求項1記載のシステム。

【請求項4】アダプタがホスト・コンピュータに内在する、請求項1記載のシステム。

【請求項5】各々がホスト・コンピュータに関連付けられ、指定されたアレイの2次制御を有する、複数の2次アダプタを含み、指定アレイを本来制御するアダプタが使用不能な場合、2次アダプタが前記指定アレイを制御する、請求項1記載のシステム。

【請求項6】指定アレイの1次制御を有するアダプタ、及び2次制御を有するアダプタが、異なるホスト・コンピュータに内在する、請求項5記載のシステム。

【請求項7】第1のホスト・コンピュータが第1のアレイを本来制御する第1のアダプタ、並びに第2のアレイを補助的に制御する第2のアダプタを有し、第2のホスト・コンピュータが前記第1のアレイを補助的に制御する第3のアダプタ、並びに前記第2のアレイを本来制御する第4のアダプタを有する、請求項6記載のシステム。

【請求項8】第1のアダプタが第1のアレイの1次制御、並びに第2のアレイの2次制御を有し、第2のアダプタが前記第2のアレイの1次制御、並びに前記第1のアレイの2次制御を有する、請求項5記載のシステム。

【請求項9】前記第1及び第2のアダプタが同一の前記ホスト・コンピュータに内在する、請求項8記載のシステム。

【請求項10】RAID機構に従い、アレイ内に記憶されるデータを分配するRAID制御装置機能を前記アダプタ内に含む、請求項1記載のシステム。

【請求項11】前記アダプタ通信相互接続がSSAインタフェースである、請求項1記載のシステム。

【請求項12】複数の記憶装置アレイ上の複数のホスト・コンピュータのデータをアクセスする方法であって、

a) 少なくとも1つのアダプタを各ホスト・コンピュータに関連付けるステップと、

b) アダプタを記憶装置アレイに1次制御装置として関連付けるステップと、

c) 全ての前記アダプタ間で通信相互接続を関連付けるステップと、d) 前記ホストに関連付けられるアダプタにデータ・アクセス要求を送信するステップと、

e) 要求データがアダプタにより本来制御される前記記憶装置アレイ内に記憶されているか否かを識別するステップと、

f) アダプタにより本来制御される記憶装置アレイ内に記憶されていないデータに対するデータ・アクセス要求を、前記通信相互接続を通じて、前記要求データを有する前記記憶装置アレイを本来制御するアダプタに送信するステップと、

を含む、方法。

【請求項13】g) アダプタを記憶装置アレイの2次制御装置として関連付けるステップと、

h) 前記記憶装置アレイを本来制御するアダプタが使用不能な場合を判断するステップと、

i) 記憶装置アレイの1次制御を有するアダプタが使用不能な場合、前記記憶装置アレイの2次制御を有する前記アダプタを当該記憶装置アレイの制御装置として使用するステップと、

を含む、請求項12記載の方法。

【請求項14】複数のホスト・コンピュータ及び複数の記憶装置アレイを含むネットワーク化コンピュータ・システムにおいて、各ホスト・コンピュータがアダプタ・カードを有し、前記アダプタ・カードが前記記憶装置アレイを本来制御し、前記システム内の全ての前記アダプタ・カード間の通信相互接続を有するものにおいて、前記アダプタ・カードと一緒に使用されるコンピュータ・プログラムであって、

複数の前記記憶装置アレイ上に記憶されるデータに対する全ての前記ホスト・コンピュータのアクセスを提供する、コンピュータ読出し可能プログラム・コード手段を有するコンピュータ使用可能媒体と、

ホスト・コンピュータに関連付けられるアダプタ・カードがデータ・アクセス要求を受信するコンピュータ読出し可能プログラム・コード手段と、

要求データがアダプタにより本来制御される前記記憶装置アレイ内に記憶されているか否かを識別するコンピュータ読出し可能プログラム・コード手段と、アダプタにより本来制御される記憶装置アレイ内に記憶されていないデータに対するデータ・アクセス要求を、前記通信相互接続を通じて、前記要求データを有する前記記憶装置アレイを本来制御する前記アダプタ・カードに送信するコンピュータ読出し可能プログラム・コード手段と、

を含む、コンピュータ・プログラム。

【請求項15】1次アダプタが故障の場合に、2次アダプタが前記1次アダプタに対応する記憶装置アレイの制御を引き継ぐことを可能にするコンピュータ読出し可能プログラム・コード手段を含む、請求項14記載のコンピュータ・プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はデータ記憶システムに関し、特に複数のホスト・システムにより共用される複数の記憶システムに関する。

【0002】

【従来の技術】コンピュータ利用の発展は、コンピュータ・システムのデータを記憶する柔軟性のある高い可用性のシステムに対する需要を増加させた。多くの企業が、パーソナル・コンピュータ及びワークステーションを含む複数のホスト・コンピュータ・システムを有し、これらは独立に機能するか、ネットワークを通じて接続される。データが全てのホスト・システムによりアクセスされ得るように、複数のホスト・システムが複数の記憶システムの共通のプールにアクセスできることが望ましい。こうした構成は、任意のホスト・システムが使用可能な総データ量を増加させる。また、ホスト間で作業負荷が配分され、システム全体が任意のホストの故障から保護され得る。

【0003】記憶システムに記憶されるデータの可用性を保護することも重要である。データ保護のための1つの機構は、RAID (Redundant Array of Independent Disks) 機能を組み込むことである。RAID技術の概念及び変形は、記憶装置業界では既知である。RAIDのレベルが、Pattersonらによる“A Case for Redundant Array of Inexpensive Disks (RAID)”, proceedings of the 1988 ACM SIGMOD Conference on Management of Data, Chicago, Illinois, June, 1988に述べられている。通常のRAIDシステムは、RAID制御装置並びにアレイとして構成される複数の記憶装置を含み、後者はディスク・ドライブとしても参照される直接アクセス記憶装置(DASD)などを含む。データはパリティ情報を用いてシステム上で保護され、パリティ情報もアレイの一部として記憶される。RAIDレベル0アレイは通常、パリティ保護を有さず、データが全てのDASDに渡りストライプされるアレイを指す。RAID1システムでは冗長保護のために、あるDASDからのデータが第2のDASD上にミラーされる。RAID5アーキテクチャでは、論理パリティ・ディスクを指定することにより、RAIDオペレーションの効率及び信頼性が向上される。この論理パリティ・ディスクは、アレイ上の各ディスクに渡り物理的にストライプされ、従って、どのディスクもアレイ全体のパリティを含まない。JBOD (just a bunch of disks) は通常、ストライピン

グまたは冗長性を有さないDASDのアレイを指す。RAID0、RAID1、RAID3及びRAID5の各レベルが他に比較してより好適と思われる、特定のオペレーションが存在する。例えばRAID5は、データに対する非常に多数の同時アクセスを要求するシステムに好適である。

【0004】更に、記憶システムが読出しキャッシュまたは書込みキャッシュのいずれかのキャッシュを含むことが望ましい。業界では、不揮発性キャッシュの提供も既知であり、このキャッシュに書き込まれたデータは、ディスク・アクセス並びにディスク自身へのデータの実際の書込みを待機する必要無しに、あたかもディスクに書き込まれたかのように見なされる。

【0005】データ・アクセスに対する性能及び高い可用性が保証されるように、ハードウェア故障から保護するための冗長バスを提供することも望ましい。

【0006】複数のホストが複数のコンピュータ・システムにアクセスすることを可能にする従来の解決策では、ホスト・アダプタ・カード、アウトボード・ディスク制御装置、及び標準ネットワーク通信システムの組み合わせを用いた。

【0007】複数のホストが複数の記憶装置アレイにアクセスすることを可能にするように、ネットワーク接続されたコンピュータ記憶システム構成の従来例が、図1及び図2に示される。図1は複数のホスト・システム10を有するシステムを示す。各ホスト・システムは、ディスク・アレイ12へのそれぞれのインタフェース11を有する。ホスト・システムは、互いにネットワーク13及びネットワーク・ファイル・システムを通じて通信する。こうしたネットワーク・ファイル・システムには、サン・マイクロシステムズ社から提供されるネットワーク・ファイル・システム(NFS)がある。あるホストが異なるホスト・システムに接続され、そのホストにより制御されるデータをアクセスする必要がある場合、データ・アクセス要求がネットワーク・サーバを通じて、データが記憶されるアレイを制御するホストに経路指定される。しかしながらこの解決策には、要求の送信が遅いことによる制限が存在する。また、ネットワーク時間の使用がこの種のオペレーションには不適切であり、代わりにホスト間の他の種類の通信のために必要とされる。また、ネットワークがこの種のデータ・アクセスに対して最適化されない。

【0008】図2は、別々の制御装置サブシステム15が複数のホスト16によりアクセス可能なシステムを示す。制御装置サブシステム15は、このサブシステムに接続されるアレイ18及び19のための制御機能を提供する。これらの機能には、RAIDのパリティ及びストライピング機能、キャッシュ読出し機能、及び不揮発性キャッシュ書込み機能が含まれる。ホスト16は共有制御装置を通じて、データへのアクセスを有する。ホスト

はいずれかの制御装置を通じて要求を送信し、制御装置はアレイをアクセスし、要求されたデータをホストに返送する。しかしながら、図2に示されるシステムは多数の制限を有する。1サブシステムに接続可能なホスト・コンピュータの数に制限があり、サブシステムにより制御されるアレイの数にも制限がある。また、図2に示される従来技術では、ホストがそれ自身のDASDセット用の独自の制御装置を有して独立しないように、ホストとアレイ間に別々のレベルの制御装置を有する。更に、アウトボード制御装置が追加の電子回路、電源、及びパッケージングを要求し、そのことがコストを増加させ、システム全体の信頼性を低下させる。

【0009】

【発明が解決しようとする課題】従って、ホスト及び記憶装置アレイのより大規模な接続性を可能にする、より安価でよりスケーラブルな解決策が必要とされる。こうしたシステムが高い可用性及び優れた性能を有することが望ましい。

【0010】

【課題を解決するための手段】本発明は従来システムにおける上述の問題を解決する一方、安価でよりスケーラブルな解決策を提供する。本発明は、ホストに内在することができ、多数の記憶装置アレイを制御できるホスト・アダプタ・カードを使用するアーキテクチャを提供する。

【0011】各記憶装置アレイ上のデータが、任意のホスト・コンピュータによりアクセスされ得るように、複数のホスト・コンピュータのデータを複数の記憶装置アレイ上に記憶するシステムが提供される。複数のアダプタ・カードが使用され、各アダプタは、指定された記憶装置アレイに対する制御装置機能を有する。システム内の全てのアダプタ間に、アダプタ通信インタフェース

(相互接続)が存在する。更に、ホスト・コンピュータ内で実行されるアプリケーション・プログラムとアダプタとの間に、ホスト・アプリケーション・インタフェースが存在する。アプリケーション・プログラムにより、ホスト・アプリケーション・インタフェースを通じて第1のアダプタに、その第1のアダプタにより本来制御されない記憶装置アレイに記憶されるデータに対するデータ要求が生成されるとき、データ要求はアダプタ通信インタフェースを通じて、要求データが記憶される記憶装置アレイを本来制御するアダプタに伝達される。

【0012】好適な態様では、各ホスト・コンピュータは1つ以上のアダプタを有し、各アダプタはRAID制御装置、1つ以上の外部インタフェース、及び読出し及び書き込みキャッシュを含む。ディスク・ドライブはRAID0、RAID1、RAID3またはRAID5などのRAID機構内に、1つ以上のアレイとして構成される。所与の時点において、各アレイは1つのアダプタにより制御され、そのアレイへの全てのアクセスがそのア

ダプタを通過する。このことは、同時にオーバーラップするパリティ更新による潜在的な衝突を容易に解決する。また、同一データの複数のコピーを複数のキャッシュに記憶することによるコヒーレンス問題を回避する。アダプタ間の相互接続は、アダプタとディスク間の通信の他に、アダプタ間(ピア・ツー・ピア)の通信を可能にする。

【0013】好適な態様では、更に各記憶装置アレイの2次制御を有する複数のアダプタが存在する。2次アダプタは、指定された記憶装置アレイを本来制御するアダプタが使用不能な場合に、その指定された記憶装置アレイを制御する。アダプタ通信インタフェースは、2次アダプタを含む全てのアダプタを相互接続する。

【0014】本発明の目的は、複数のホストが複数のアレイをアクセスすることを可能にする高い可用性のシステムを提供することである。シリアル・ストレージ・アーキテクチャ(SA: Serial Storage Architecture)のループ・トポロジを用いる技法が、128のホスト・システムの相互接続を可能にする。

【0015】ホスト・アダプタ・カードは、パーソナル・コンピュータ及びワークステーション用のマイクロチャネル・バスまたはPCIバスなどの、多くのホスト・バスと互換性を有し、このことは安価でよりスケーラブルな解決策を提供する。SAインタフェースに加え、ファイバ・チャネル・インタフェース、FC-AL(Fiber Channel-Arbitrated Loop)も、SCSI2などの既存の並列インタフェースと共に使用され得る。

【0016】

【発明の実施の形態】図3は、本発明の機能及びオペレーションの主要要素の論理概要を示す。各ホスト・コンピュータ20は、1つ以上のアダプタ22を有する。各アダプタはRAID制御装置、1つ以上の外部インタフェース、並びに読出し及び書き込みキャッシュを含む。ホスト・コンピュータは、ディスク・ドライブのアレイの共通プールを共有する。ディスク・ドライブはRAID0、RAID1、RAID3またはRAID5、或いはJBODなどの1つ以上のアレイに構成される。これらは当業者には既知のシステムである。任意の時点において、各アレイは1つのアダプタにより制御され、アレイへの全てのアクセスがそのアダプタを通過する。このことは、同時にオーバーラップするパリティ更新による潜在的な衝突を容易に解決する。また、同一データの複数のコピーを複数のキャッシュに記憶することによるコヒーレンス問題を回避する。各アレイは1アダプタによってのみ制御されるので、あるアレイの制御アダプタ以外のアダプタから発信されるそのアレイに対する要求は、最初に制御アダプタに経路指定されなければならない。その後、制御アダプタがアレイをアクセスし、結果を発信元リクエストに返却する。アダプタ22は、アダプタとディスク間の通信の他に、アダプタ間の通信を可能にする

相互接続23を通じて相互接続される。

【0017】ホスト・システムの例として、IBM AIXオペレーティング・システムを実行するIBM Risc System/6000マシンが挙げられる。当業者には既知の多くの他のホスト・システムも使用され得る。

【0018】図3及び図4を参照して、ホスト1がアダプタBにより制御されるアレイ2からデータを読み出した状況では(ブロック24)、ホスト1はI/O要求をアダプタAに発行する(ブロック25)。アダプタAはレジストリを調査し、アレイ2がアダプタBにより制御されることを判断する(ブロック26)。発信元アダプタはI/O要求をアダプタBに転送する(ブロック27)。アダプタBはそのキャッシュを探索し、必要に応じてディスクをアクセスすることにより、I/O要求を実行する。アダプタBは次に、読み出されたデータをアダプタAに返却する(ブロック28)。アダプタAは読み出されたデータをホスト・メモリに記憶し、I/O要求が完了したことを示すために、ホストに割込みを発生する。

【0019】帯域幅、ファン・アウト、及び当業者には既知の要求される可用性に依存して、多くの技法が相互接続性23を提供するために使用され得る。例えば、1つ以上の並列バス、直列ループ、または直列スイッチが存在し得る。本発明の好適な態様は、相互接続アーキテクチャとしてSSAに関連して述べられるが、他のアーキテクチャも使用可能である。

【0020】SSAはANSI (American National Standard Institute: 米国規格協会) により開発されている。X3T10.1. SSAは、特にディスク・ドライブ、テープ・ドライブ、CD-ROM、光ドライブ、プリンタ、スキャナ、及び他の周辺装置などのI/O装置を、ワークステーション・サーバ(ホスト・システム)及び記憶サブシステムに接続するために設計された直列インタフェースである。当業者にはSSAアーキテクチャは既知であるので、ここではそのアーキテクチャ及びそのオペレーションの詳細には触れないことにする。SSAの詳細については、"Information Technology - Serial Storage Architecture - Transport Layer 1 (SSA-TL1)", ANSI X3T10.1/0989D", "Information Technology - Serial Storage Architecture - Physical 1 (SSA-PH1)", ANSI X3T10.1/xxxD", 及び"Information Technology - Serial Storage Architecture - SCSI-2 Protocol 1 (SSA-S2P)", ANSI X3T10.1/1121D"を参照されたい。

【0021】リンクまたはバスは、データを銅線などの伝送媒体上において、信号として伝送するために使用されるコネクタを指す。直列リンクは単一信号サイズのケーブルを使用することができ、そこでは伝送データ単位が通信バス上で直列化される。直列コネクタは通常、双方向(全2重)通信を有する。(ディスク・アレイに関して) ループ構成において使用される場合、ループ内の

各ドライブへの帯域幅を倍増するために、或いは接続の故障に備え、ドライブへの別の経路を提供するために、第2のバスが使用され得る。SSAは、全2重通信を提供する2信号接続(送信及び受信)を提供する。

【0022】SSAは、直列に接続される周辺装置をアドレス指定するSCSI (small computer system interface) 仕様の論理面を使用する。これらのSCSI仕様は、SSAの物理仕様にマップされる。すなわち、SSAは様々な上位レベル・プロトコルにおける、特に記憶装置アプリケーション用のSCSI-2における、トランスポート層として使用され得る。SSA上のSCSI-2は、SCSI規格において定義されるものと類似のアドレス機構を保持する。SCSI規格では、イニシエータ、ターゲット、及び論理単位が存在する。

【0023】大部分の基本SSAネットワークは、単一ポート周辺装置に接続される単一ポート・ホスト・アダプタを含む。直列接続は、情報フレームを伝達するために使用される4ワイヤーを含む。4ラインは、プラス/マイナス・ライン出力(送信)及びプラス/マイナス・ライン入力(受信)を含む。"ポート"は、リンク(1送信バス及び1受信バス)の1端をサポートするハードウェア及びファームウェアを含むゲートウェイを指す。あるノード内のポートは、別のノード上のポートにリシクを介して接続される。

【0024】SSAにおけるポートは、2つの20メガバイト/秒の対話を同時に実行することができ、一方はインバウンドで他方はアウトバウンドである。ループ内の各リンクは独立に動作するので、総ループ帯域幅は単一のリンクよりも高くなり得る。ループはまた単一の故障を許容し得る。なぜなら、メッセージ及びデータが時計回りまたは反時計回りに転送され得るからである。SSAデュアル・ポート・ノードは、80メガバイト/秒の総帯域幅において、4つの同時対話を伝搬できる。

【0025】ノードは、1つ以上の直列ポートを有するシステム制御装置、ホスト、または周辺装置を指す。各ノードはその特有の責任またはタスクに相当する機能を有する。イニシエータは、実行される必要のあるタスク、並びに所望のタスクを実行するターゲットを決定するノード内の機能である。ノードは1つ以上のポートを実現する。

【0026】フレームは、SSAネットワーク内の2つのポート間の情報伝送の基本単位である。フレームは、制御バイト、最大6バイトのアドレス、最大128バイトのデータ、及び4バイトのエラー検出を含む期待フォーマットを有する。ノードはフレームをポート間で経路指定することができる。ノード機能はフレームを発信または送信し得る。SSAプロトコルは特殊な文字を用いて、ノード間で伝送されるフレームの流れを調整し、フレームを確認する。フレーム多重化機能は、電信システムが同一ケーブル上で、2つ以上のメッセージをいづれ

の方向にも同時に送信できることを意味する。

【0027】SSAは、ストリング、ループ及びスイッチ構成を含む複数のトポロジにより実現され得る。通常の単一ループ構成では、128デュアル・ポート・ノード（周辺装置またはホスト）がサポートされ得る。複雑なスイッチ構成では、ノードの理論的な最大数は200万以上である。ループ・トポロジはネットワーク内の各ノードへの代替パスを許可し、ネットワークを1箇所の故障として除去する。

【0028】ゲートウェイは2つのノード間で確立され、SSAネットワーク上で全2重通信を提供する。ノードは、例えばディスク・アクセスなどの機能を実行するために、別のノードにトランザクションを発行する。ゲートウェイは2つの接続を含み、各々がそれぞれの方向を担う。マスタ（トランザクションを発行する側）がマスタ制御ブロックを生成する。ゲートウェイはトランザクションをフレーム形式で、ネットワーク上に送信する。スレーブ側のゲートウェイは、トランザクション・フレームを受信し、アドレス指定サービスを呼び出すタスク制御ブロックを生成する。

【0029】図5はアダプタ・カード・ハードウェアのブロック図を示す。アダプタはマイクロプロセッサ30を有し、これはマイクロプロセッサ・バス32を介して、RAM33、不揮発性RAM34、及びROMブートストラップ35に接続される。RAM33は、必要なコード及び制御ブロック（図6でより詳細に述べられる）を含む。マイクロプロセッサ・ブリッジ36は、マイクロプロセッサ30をローカル・バス37を介して、不揮発性キャッシュ38及び読出しキャッシュ40に接続する。XORハードウェア42もまた、RAIDパリティ計算を実行するために提供される。ホスト・ブリッジ44は、マイクロチャネルまたはPCIバス45などのホスト・インタフェースを通じて、主ホスト・ハードウェアへの接続を提供する。2つのSSAデュアル・ポート・チップ46及び47が存在し、一方は他のアダプタとの相互接続用であり、他方は記憶装置アレイとの相互接続用である。各ブロックは別々のASIC（アプリケーション特有の集積回路）であり得る。XOR機能は、読出しキャッシュ制御装置と1ASIC上に結合され得る。

【0030】図6は本発明を実現するために、ホスト及びアダプタ・カード上で実行されるソフトウェアのブロック図を示す。ホスト・システムは中央処理ユニット（CPU）（図示せず）、及び実行中にソフトウェアが記憶されるRAM（図示せず）を有する。アプリケーション・プログラム56はRAM内で実行され、ファイル・システム57に、またはオペレーティング・システム60を通じて直接デバイス・ドライバ58に、I/O要求を生成する。デバイス・ドライバは、オペレーティング・システムとのインタフェース61、ソフトウェア・

バス62、及び、アダプタ・カードへのゲートウェイ63を有する。

【0031】アダプタ・カード64は独自のマイクロプロセッサ及び独自のRAMを有し、RAMは実行の間に、本発明を実現するソフトウェアを含む（アダプタ内のハードウェアの詳細については図5を参照）。アダプタ・カードは、デバイス・ドライバと対話するゲートウェイ66、並びにキャッシュ制御装置68及びRAID制御装置70と対話するソフトウェア・バス67を有する。アダプタ・カード64は更にレジストリ72を有し、これは本来そのアダプタ・カードにより制御される記憶装置のアレイ、並びに他のアダプタ・カードを介してアクセス可能なアレイを識別する。ソフトウェア・バスはゲートウェイ74とも対話し、ゲートウェイ74はインタフェース・チップを通じて、他のアダプタ・カードとのピア・ツー・ピア・リンク75を可能にする。ディスク・インタフェース76はディスク・アレイ77と対話する。

【0032】アレイへの接続性を増強するために、各アレイは1次及び2次（バックアップ）制御アダプタを提供される。ループ内の全てのディスクが両方のアダプタに接続されるので、両方のアダプタが機能しているときには、他のアダプタにより本来制御されるアレイをアクセスする要求は、最初にピア・ツー・ピア・リンク75を通じて、他のアダプタに経路指定される。結果が再度ピア・ツー・ピア・リンク75を通じて要求元アダプタに返送され、そこからホストに返却される。

【0033】1つの構成では、複数のディスク・アレイが同一のSSAディスク・アレイ・ループ上に存在する場合、1アダプタが、あるアレイの1次制御アダプタであり、別のアレイの2次制御アダプタでもある。図7を参照すると、各ホスト80乃至85は1つのアダプタ・カード86乃至91を有し、これらのアダプタ・カードはディスク・アレイ92乃至97の1次または2次アダプタとして作用する。6つのホスト・コンピュータだけしか図示されていないが、前述のように、より多くのホストが相互接続され得る。各ディスク・アレイ92乃至97は、そのアレイ用の制御装置として作用する1次及び2次アダプタを有する。例えば、アダプタA1（86）は、ディスク・アレイA（92）の1次制御装置であり、ディスク・アレイB（93）の2次アダプタでもある。アダプタA2（87）は、ディスク・アレイB（93）の1次制御装置であり、ディスク・アレイA（92）の2次アダプタでもある。アダプタは全てSSAループ98を通じて相互接続される。ディスク・アレイA及びディスク・アレイBなどのディスク・アレイ対は、SSAデバイス・ループ99を通じて接続される。1次アダプタがアクティブの時、I/O要求は1次アダプタを通じて転送される。1次アダプタ・カードが故障の場合にも、他のホストは2次アダプタを通じて、ディ

スク・アレイをアクセスすることができる。

【0034】システムは、各ホスト・コンピュータ内の複数のアダプタ、並びにループまたはデュアル・スイッチなどの相互接続による代替パスを含むことにより、1箇所の故障に対処可能に構成され得る。図8は3つのSSAループを実現する高可用性の構成を示す。各アダプタ・カード110は、2つのデュアル・ポートSSAノード112、113を含む。一方のノード112は、ホスト・ループ114に相当する外部ループに接続される。このループは、発信元アダプタと1次アダプタ間の通信のためにのみ使用される。他のノード113は、デバイス・ループ116に相当する2つの内部ループの一方に接続される。各デバイス・ループは、1次アダプタと、対応するアレイ内のディスクとの間の通信を提供する。このループは1次アダプタと2次アダプタ間の通信にも使用される。

【0035】図8を参照すると、アダプタAはアレイ1の1次制御を有し、アダプタDはアレイ1の2次制御を有し、アダプタBはアレイ2の2次制御を有し、アダプタCはアレイ2の1次制御を有する。ホスト1によりアレイ1に対して生成されるI/O要求は、アダプタAに直接発行される。アダプタAは次にこれらをディスク読出し/書込みコマンドに変換する。ホスト1によりアレイ2に対して生成されるI/O要求は、アダプタB及びリンク2を介して、アダプタCに発行される。アダプタCは次にこれらをディスク読出し/書込みコマンドに変換する。同様に、ホスト2によりアレイ2に対して生成されるI/O要求は、アダプタCに直接発行される。アダプタCは次にこれらをディスク読出し/書込みコマンドに変換する。ホスト2によりアレイ1に対して生成されるI/O要求は、アダプタD及びリンク4を介して、アダプタAに発行される。アダプタAは次にこれらをディスク読出し/書込みコマンドに変換する。図示の例では、ディスクは分散パリティを有する7+P（7データ・ディスクと1パリティ・ディスク）構成などの、2つのRAID5アレイとして構成される。ディスクは故障許容能力及び冷却機構を具備して、外部エンクロージャ内に実装される。

【0036】代わりに、ディスクが8つのRAID1アレイとして構成されてもよい。この場合、各アレイの1ディスクは、各ホスト・コンピュータ内に実装され得る。また、複数のアレイが同一デバイス・ループ内に存在する場合、アダプタが1アレイの1次制御アダプタとして、かつ第2のアレイの2次制御アダプタとして作用することができる。例えば、アダプタAとアダプタDとの間のデバイス・ループ内に、2つの3+P RAID5アレイが接続される場合、アダプタAは、第1の3+P RAID5アレイを形成する上方ドライブ・グループ120の1次制御、並びに第2の3+P RAID5アレイを形成する下方ドライブ・グループ122の2次

制御を有する。また、アダプタDは下方グループ122の1次制御、並びに上方グループ120の2次制御を有する。また、ホストは3つ以上のアダプタを有することができる。

【0037】あるアダプタがダウン（故障）すると、そのオペレーションは2次アダプタにより引き継がれる。従って、ホストは依然として、その別のアダプタを通じて、任意のアレイをアクセスすることができる。

【0038】あるアダプタがダウンすると、そのアダプタ上で進行中のあらゆる遠隔I/O要求が、発信元アダプタ・カード内のSSAゲートウェイにより、エラー終了される。トランザクションを転送したモジュールは、2次アダプタがダウンしたアレイを引き継いだことを同報するのを期待して、数秒を待機する。次に、失敗したトランザクションが2次アダプタに再送され得る。

【0039】SSAアドレス指定機能は、ホスト・ループを最大128アダプタにまで拡張可能である。多数のアダプタを有すると、ホスト・ループが究極的な帯域幅制限となる。なぜなら、ホスト・ループがほぼ全てのI/O要求により共有されるからである。平均的に要求のランダム配布を想定すると、各要求はループに沿って約半周分伝搬されなければならない、使用可能な総帯域幅は 2×20 、すなわち40メガバイト/秒の全2重となる。可能なデッドロックを回避するために、SSAフレームはループの1リンクまたは1ノードを通じて経路指定されてはならない。読出し対書込み比率が1対1の場合、総I/O帯域幅は80メガバイト/秒であり、オーバーヘッドを無視する。ホスト・ループをマルチポートSSAスイッチにより置換することにより、より高い帯域幅が達成され得る。

【0040】図8に示される構成では、任意の1アダプタ・カード上で受信されるトランザクションの幾つかだけが、そのカード上のRAIDソフトウェアにより処理される。遠隔資源に向けられるトランザクションは、SSAゲートウェイを通じて経路指定される。SSAゲートウェイは、自身が見出す各遠隔アダプタをアドレス指定するためのノード番号を提供する。

【0041】アレイの一部に相当する各ディスクが、制御アダプタに割当てられる。これはそのディスクの1次アダプタとして認識される。各ディスクは、2つのアダプタだけを有するループ内に置かれる。各アダプタはノード番号を割当てられる。この番号はホストにより提供され、固有の識別子として見なされる。アレイが生成されるとき、ディスクはそのアレイを生成するために使用されるアダプタのノード番号を、自動的にマーク付けされる。ディスクは更に、ループ内の他のアダプタのノード番号によってもマーク付けられる。他のアダプタはアレイのバックアップ・アダプタであり、2次アダプタとして参照される。通常のオペレーションでは、1次アダプタは、アレイ内のディスクを制御するRAIDコード

を実行する。2次アダプタは後述されるように、1次アダプタにより指示される書き込みキャッシュ及び不揮発性メモリの2重コピーを保持するだけである。故障を検出すると、2次アダプタは1次アダプタのオペレーションを引き継ぐ。

【0042】各アダプタ・カードはレジストリを含み、これはソフトウェア・バスを介してアクセス可能な中央サービスである。レジストリは全てのアダプタのリスト、並びにシステム内の全てのアレイのリストを保持する。図9を参照すると、アダプタ・リスト130内の各エントリはノード番号131を含む。

【0043】アレイ・リスト133内の各エントリは、そのアレイを現在制御しているアダプタ（1次または2次アダプタ）の資源ID134、及びノード番号135を含む。資源IDは、デバイス・ドライバが読み出し及び書き込みのために、アレイをオープンするために使用する。資源IDは、タイプ・フィールド137及び資源番号138を含む。タイプ・フィールドは、例えばRAIDレベル0、1、3または5などの、アレイ・タイプを指定する。資源番号は、アレイが構成されるときに割当てられる固有の識別子である。

【0044】各DASDは構成レコード140を記憶し、これは親アレイが構成されるときに生成される。構成レコードは資源ID141、アレイ・パラメータ142（例えばストライプ・サイズ）、アレイ内の他のDASDの通し番号143、1次アダプタ・ノード144、及び2次アダプタ・ノード145を記憶する。これら全てはホスト構成ユーティリティを用いて、手動式に割当てられる。構成レコードは更に、アレイを現在制御しているのが1次アダプタか、それとも2次アダプタかを示すフラグ146を含む。このフラグは前述のように、1次及び2次アダプタ内のレジストリにより管理される。

【0045】レジストリの構成時、アダプタは、アレイの一部に相当する各ディスク上の1次ノード番号を調査し、それが自身に帰属するか否かを判断し、ディスクをRAIDファームウェアに渡す。アダプタが特定のディスクの1次制御装置でない場合、アダプタは他のアダプタをピング（ping）し（信号を送り）、それが動作しているか否かを判断する。他のアダプタが動作していない場合、最初のアダプタがそのディスクのアレイ制御装置として引き継ぐ。これは“フェール・オーバー（fail-over）”として参照される。レジストリがディスクを、あるシステムから別のシステムにフェール・オーバーさせる。フェール・オーバーは、あるアダプタが他のアダプタが作動状態から故障状態に移移したことを確認するとき、発生する。これが発生するとき、フェール・オーバーが実行される。全てのディスクのスイッチ・オーバー（切り替え）が、例えばSCSI予約コマンドによりロックされる。フラグが次に各ディスク上において変更され、ディスクが1次アダプタではなく2次アダプタにより保守さ

れることを示す。全てのディスクが次に解放される。ディスクの1つが予約され得ない場合、これまでに予約された全てのディスクから、予約が解放される。プロセスがランダム期間の後、バック・オフ（撤回）され、再試行される。予約期間の後、依然としてエラーが発生する場合には、全てのフラグが1次状態に戻され、プロセス自身はバック・オフされ、再試行される。このプロセスの間に電源が故障した場合、電源の回復時に、両方のアダプタが不一致を認識し、デフォルト指定により、全てのディスクを1次アダプタにセットする。

【0046】“フェール・バック（fail-back）”は、前述のフェール・オーバー・プロシジャの逆である。これは両方のアダプタ・カードが作用しているときに発生する。プロセスは、1次アダプタを周期的にピング（信号を送る）して、それが回復したかどうかを判断する2次アダプタにより開始される。ピングが成功の場合、2次アダプタ上のレジストリがそのRAID制御装置に、1次アダプタへの回復を希望する各ディスクを通知し、そのディスクが解放されるべきであることを指定する。2次アダプタ内のRAID制御装置は、最初に、指定されたディスクを含むアレイをクローズし、次にディスク自身をクローズする。アレイ内の全てのディスクがクローズされると、2次アダプタ内のレジストリが前述のようにそれらのディスクをロックし、それらが再度1次アダプタにより制御されるように、フラグをスイッチ・バックする。2次アダプタが次にディスクを解放し、1次アダプタ内のレジストリに、フェール・バック・プロセスが完了したことを伝える。

【0047】システムが不必要なフェール・オーバーの発生無しにパワー・オフされるように、システムは最初に自身がオープンした全てのアレイをクローズし、アダプタ・カードのレジストリに、クローズ・ダウンを伝える。次に1次アダプタ上のレジストリが2次アダプタ上のレジストリに、自身がクローズ・ダウンされることを伝える。これにより、2次アダプタ上のレジストリはフラグをリセットし、1次アダプタが作用していることを確認したことを示す。フェール・オーバーは、2次アダプタが1次アダプタの作動状態から故障状態への遷移を確認したときのみ発生するので、1次アダプタの続くパワー・ダウンはフェール・オーバーを引き起こさない。

【0048】ディスク・アレイは1次及び2次アダプタ間のSSAループ内に構成される。ディスク間またはディスクとアダプタ間でリンクが故障する場合、全てのディスクは依然として両方のアダプタによりアクセス可能である。なぜなら、伝送がループを通じて、両方向に経路指定され得るからである。従って、図7を参照して、リンク101が故障しても、アダプタA1（86）は依然、リンク102からリンク104を通じて、更にアダプタ・カードA2のインタフェース・チップを介して、リンク103を通じて伝送を送信することにより、アレ

イA(92)に達することができる。アダプタ・カードA2へのこの伝送は、決してアダプタ・カードA2内のソフトウェアにより処理されることはない。SSAチップは、メッセージがループ内の他のアレイに対するフレームを送信し、要求がそのアレイに送信されるべきことを検出する。これはアレイB(93)内のSSAチップにもあてはまる。

【0049】上述のプロセスは、“カット・スルー・ルーティング(cut-through routing)”として参照される。これはSSAトランスポート層内の標準機能である。カット・スルー・ルーティングは、発信元アダプタからアレイの制御アダプタにデータ・アクセス要求を転送するために使用される。各デュアル・ポート・ノード(アダプタまたはディスク・ドライブ)は、SSAチップ内にハードウェア・ルータを有する。ルータはアドレス・フィールドの第1バイトを検査して、フレームをループ内の次のノードに転送すべきか否かを判断する。発信元アダプタは、要求をループ上に送信するときに、1次アダプタのバス・アドレスをフレーム・アドレス・フィールドに配置するだけでよい。

【0050】バスはシステム・パワーオン時に決定される。イニシエータ・アダプタはネットワークを“ウォーク(walk)”して、構成を判断し、各ノードのエントリを有する構成テーブルを作成する。各エントリはそのノードのバス・アドレスを含む。代替バスが存在する場合、イニシエータは一般に、最も少ないリンクを使用するバスを選択する。これはリンクが切断されると、変化し得る。

【0051】デバイス・ドライバは、要求データを含むアレイを識別する。デバイスが次にアダプタ上のレジストリに問い合わせ、局所的に制御されるアレイを決定する。ホストが複数のアダプタを有する場合、デバイス・ドライバは要求を要求データと一緒に、本来アレイを制御するアダプタに送信する(但し、アレイが局所的に制御される場合)。それ以外では、遠隔アレイに対して、デバイス・ドライバは、各アダプタを介してアレイの半分をアドレス指定するか、各アダプタを介して代替要求を送信することにより、アダプタ間で負荷を均等化しようとする。

【0052】不揮発性RAMの内容は、1次アダプタにより、2次アダプタ内の不揮発性RAM内にミラーされる。RAIDモジュールはこれを、2次アダプタ・カード上のSSAゲートウェイ及び不揮発性RAMマネージャを介して実行する。

【0053】通常オペレーションでは、1次アダプタはデバイス・ループを用いて、その書き込みキャッシュの2重コピーを2次アダプタ内に保持し、更新されるアレイ領域を示すメタ・データに対しても同様である。このことは書き込みキャッシュ内のデータを解放したり、アレイを腐敗すること無く、1次アダプタから2次アダプタへ

のフェール・オーバーを可能にする。

【0054】1次アダプタが書き込みコマンドを実行するとき、1次アダプタはデータのコピーを2次アダプタ内の書き込みキャッシュに送信する。1次アダプタがデータを書込みキャッシュからディスクにデステージ(destage)するとき、1次アダプタは更新されるアレイ領域を示すメッセージを、2次アダプタに送信する。双方のアダプタは通常、この情報を不揮発性メモリに記憶し、デステージが完了する以前の電源障害から保護する。1次アダプタがデータをデステージし、パリティを更新すると、1次アダプタは第2のメッセージを2次アダプタに送信する。2次アダプタは次に、その書き込みキャッシュ及び不揮発性メモリから、対応するレコードを削除する。

【0055】本発明は特に好適な態様に関連して述べられてきたが、当業者には、本発明がその趣旨及び範囲から逸脱すること無しに、その形態及び詳細における様々な変更が可能であることが理解されよう。

【0056】まとめとして、本発明の構成に関して以下の事項を開示する。

【0057】(1)複数のホスト・コンピュータのデータを複数の記憶装置アレイ上に記憶し、任意の記憶装置上のデータが任意のホスト・コンピュータによりアクセスされ得るようにするシステムであって、各々がホスト・コンピュータに関連付けられ、指定されたアレイの1次制御を有する、複数のアダプタと、前記アダプタ間のピア・ツー・ピア通信のためのアダプタ通信相互接続と、を含み、ホスト・コンピュータから関連アダプタへの、前記アダプタにより本来制御されないアレイに対するデータ・アクセス要求が、前記アダプタ通信相互接続を介して、前記アレイの1次制御を有する前記アダプタに伝達される、システム。

(2)前記アダプタにより本来制御される前記記憶装置を示す識別子を、当該各アダプタ内に記憶する、前記

(1)記載のシステム。

(3)前記記憶装置アレイの1次制御を有するアダプタを識別する識別子を、当該各記憶装置アレイに記憶する、前記(1)記載のシステム。

(4)アダプタがホスト・コンピュータに内在する、前記(1)記載のシステム。

(5)各々がホスト・コンピュータに関連付けられ、指定されたアレイの2次制御を有する、複数の2次アダプタを含み、指定アレイを本来制御するアダプタが使用不能な場合、2次アダプタが前記指定アレイを制御する、前記(1)記載のシステム。

(6)指定アレイの1次制御を有するアダプタ、及び2次制御を有するアダプタが、異なるホスト・コンピュータに内在する、前記(5)記載のシステム。

(7)第1のホスト・コンピュータが第1のアレイを本来制御する第1のアダプタ、並びに第2のアレイを補助

的に制御する第2のアダプタを有し、第2のホスト・コンピュータが前記第1のアレイを補助的に制御する第3のアダプタ、並びに前記第2のアレイを本来制御する第4のアダプタを有する、前記(6)記載のシステム。

(8) 第1のアダプタが第1のアレイの1次制御、並びに第2のアレイの2次制御を有し、第2のアダプタが前記第2のアレイの1次制御、並びに前記第1のアレイの2次制御を有する、前記(5)記載のシステム。

(9) 前記第1及び第2のアダプタが同一の前記ホスト・コンピュータに内在する、前記(8)記載のシステム。

(10) RAID機構に従い、アレイ内に記憶されるデータを分配するRAID制御装置機能を前記アダプタ内に含む、前記(1)記載のシステム。

(11) 前記アダプタ通信相互接続がSSAインタフェースである、前記(1)記載のシステム。

(12) 複数の記憶装置アレイ上の複数のホスト・コンピュータのデータをアクセスする方法であって、

a) 少なくとも1つのアダプタを各ホスト・コンピュータに関連付けるステップと、

b) アダプタを記憶装置アレイに1次制御装置として関連付けるステップと、

c) 全ての前記アダプタ間で通信相互接続を関連付けるステップと、

d) 前記ホストに関連付けられるアダプタにデータ・アクセス要求を送信するステップと、

e) 要求データがアダプタにより本来制御される前記記憶装置アレイ内に記憶されているか否かを識別するステップと、

f) アダプタにより本来制御される記憶装置アレイ内に記憶されていないデータに対するデータ・アクセス要求を、前記通信相互接続を通じて、前記要求データを有する前記記憶装置アレイを本来制御するアダプタに送信するステップと、を含む、方法。

(13) g) アダプタを記憶装置アレイの2次制御装置として関連付けるステップと、

h) 前記記憶装置アレイを本来制御するアダプタが使用不能な場合を判断するステップと、i) 記憶装置アレイの1次制御を有するアダプタが使用不能な場合、前記記憶装置アレイの2次制御を有する前記アダプタを当該記憶装置アレイの制御装置として使用するステップと、を含む、前記(12)記載の方法。

(14) 複数のホスト・コンピュータ及び複数の記憶装置アレイを含むネットワーク化コンピュータ・システムにおいて、各ホスト・コンピュータがアダプタ・カードを有し、前記アダプタ・カードが前記記憶装置アレイを本来制御し、前記システム内の全ての前記アダプタ・カード間の通信相互接続を有するものにおいて、前記アダプタ・カードと一緒に使用されるコンピュータ・プログラムであって、複数の前記記憶装置アレイ上に記憶され

るデータに対する全ての前記ホスト・コンピュータのアクセスを提供する、コンピュータ読出し可能プログラム・コード手段を有するコンピュータ使用可能媒体と、ホスト・コンピュータに関連付けられるアダプタ・カードがデータ・アクセス要求を受信するコンピュータ読出し可能プログラム・コード手段と、要求データがアダプタにより本来制御される前記記憶装置アレイ内に記憶されているか否かを識別するコンピュータ読出し可能プログラム・コード手段と、アダプタにより本来制御される記憶装置アレイ内に記憶されていないデータに対するデータ・アクセス要求を、前記通信相互接続を通じて、前記要求データを有する前記記憶装置アレイを本来制御する前記アダプタ・カードに送信するコンピュータ読出し可能プログラム・コード手段と、を含む、コンピュータ・プログラム。

(15) 1次アダプタが故障の場合に、2次アダプタが前記1次アダプタに対応する記憶装置アレイの制御を引き継ぐことを可能にするコンピュータ読出し可能プログラム・コード手段を含む、前記(14)記載のコンピュータ・プログラム。

【図面の簡単な説明】

【図1】複数のアレイに対する複数のホスト・アクセスを提供する従来システムのブロック図である。

【図2】複数のアレイに対する複数のホスト・アクセスを提供する第2の従来システムのブロック図である。

【図3】本発明の機能及びオペレーションの論理概要のブロック図である。

【図4】本発明を実現する方法のフローチャートを示す図である。

【図5】本発明の実現するアダプタ・カードのハードウェアのブロック図である。

【図6】本発明を実現するソフトウェアを示すブロック図である。

【図7】本発明の第1の技法を示すブロック図である。

【図8】本発明の第2の技法を示すブロック図である。

【図9】アダプタのレジストリ内及び記憶装置内に記憶されるテーブルを示す図である。

【符号の説明】

20 ホスト・コンピュータ

22 アダプタ

23 相互接続

30 マイクロプロセッサ

32 マイクロプロセッサ・バス

33 RAM

34 不揮発性RAM

35 ROMブートストラップ

36 マイクロプロセッサ・ブリッジ

37 ローカル・バス

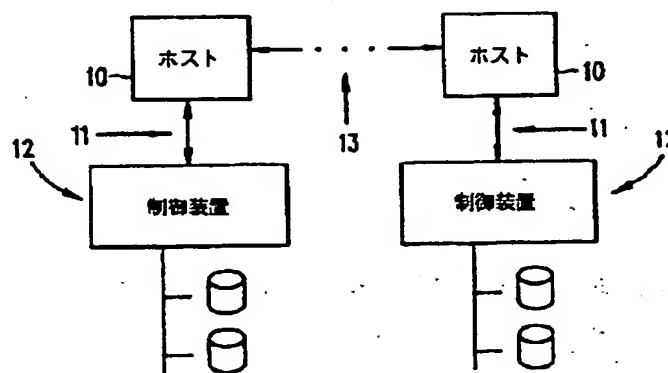
38 不揮発性キャッシュ

40 読出しキャッシュ

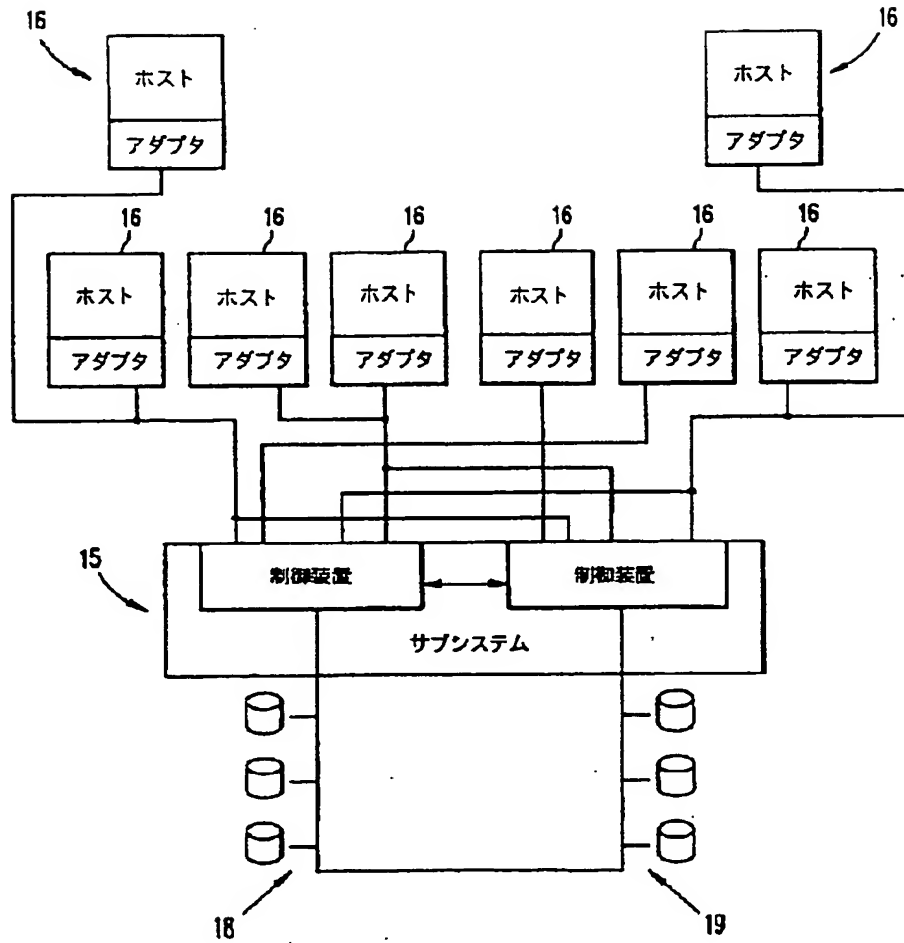
42 XORハードウェア
 44 ホスト・ブリッジ
 45 PCIバス
 46、47 SSAデュアル・ポート・チップ
 56 アプリケーション・プログラム
 57 ファイル・システム
 58 デバイス・ドライバ
 60 オペレーティング・システム
 61 インタフェース
 62、67 ソフトウェア・バス
 63、66、74 ゲートウェイ
 64、88、89、90、91 アダプタ・カード
 68 キャッシュ制御装置
 70 RAID制御装置
 72 レジストリ
 75 ピア・ツー・ピア・リンク
 76 ディスク・インタフェース
 77、94、95、96、97 ディスク・アレイ
 80、81、82、83、84、85 ホスト
 86 アダプタA1
 87 アダプタA2
 92 ディスク・アレイA
 93 ディスク・アレイB

98 SSAループ
 99 SSAデバイス・ループ
 101、102、103、104 リンク
 110 アダプタ・カード
 112、113 デュアル・ポートSSAノード
 114 ホスト・ループ
 116 デバイス・ループ
 120 上方ドライブ・グループ
 122 下方ドライブ・グループ
 130 アダプタ・リスト
 131、135 ノード番号
 133 アレイ・リスト
 134 資源ID
 137 タイプ・フィールド
 138 資源番号
 140 構成レコード
 141 資源ID
 142 アレイ・パラメータ
 143 DASDの通し番号
 144 1次アダプタ・ノード
 145 2次アダプタ・ノード
 146 フラグ

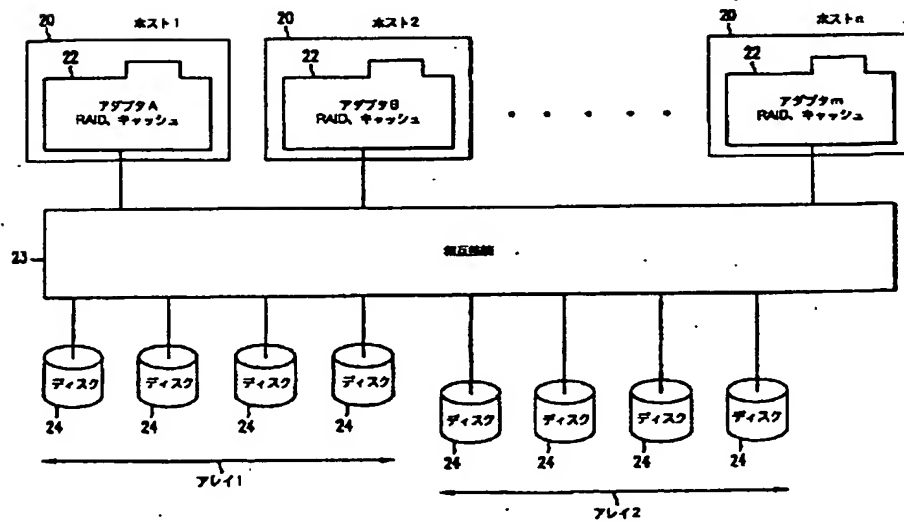
【図1】



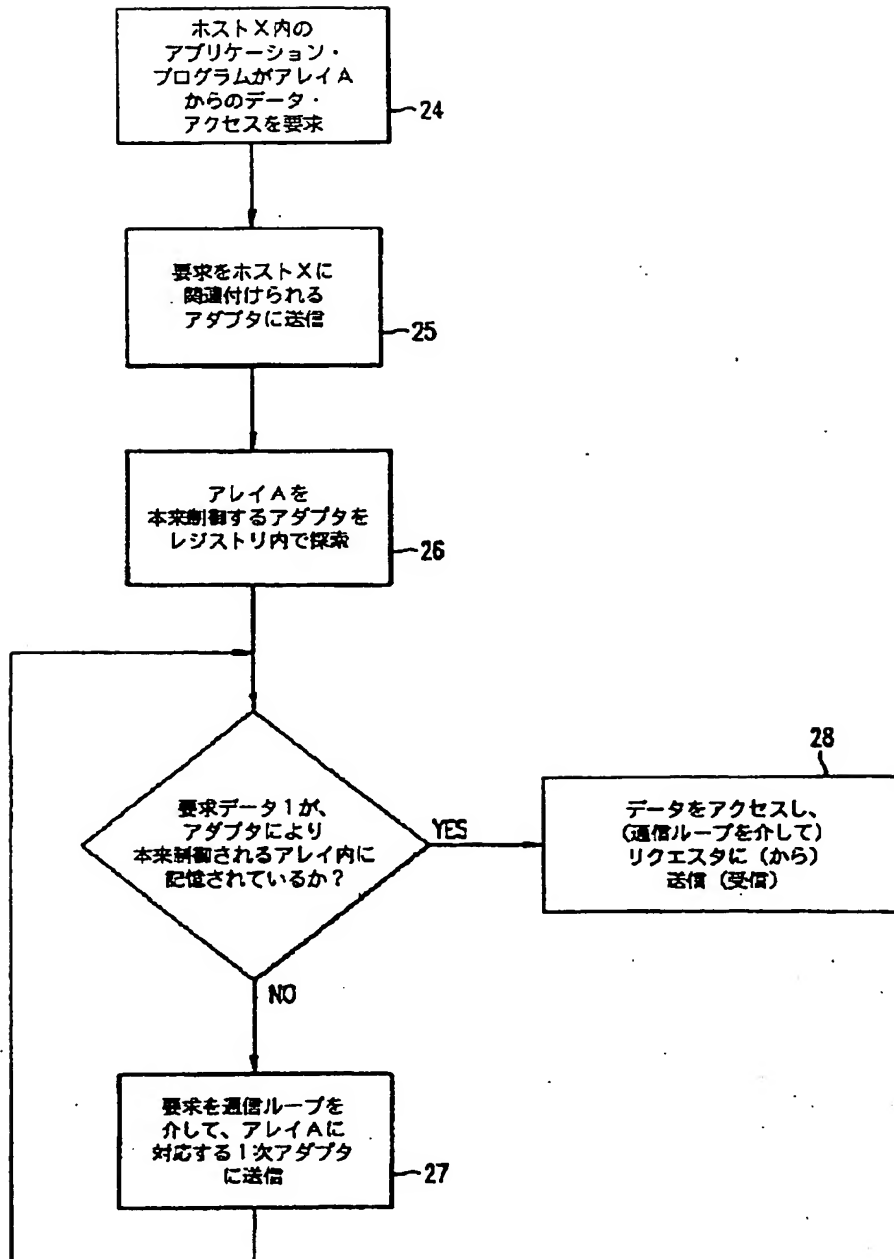
【図2】



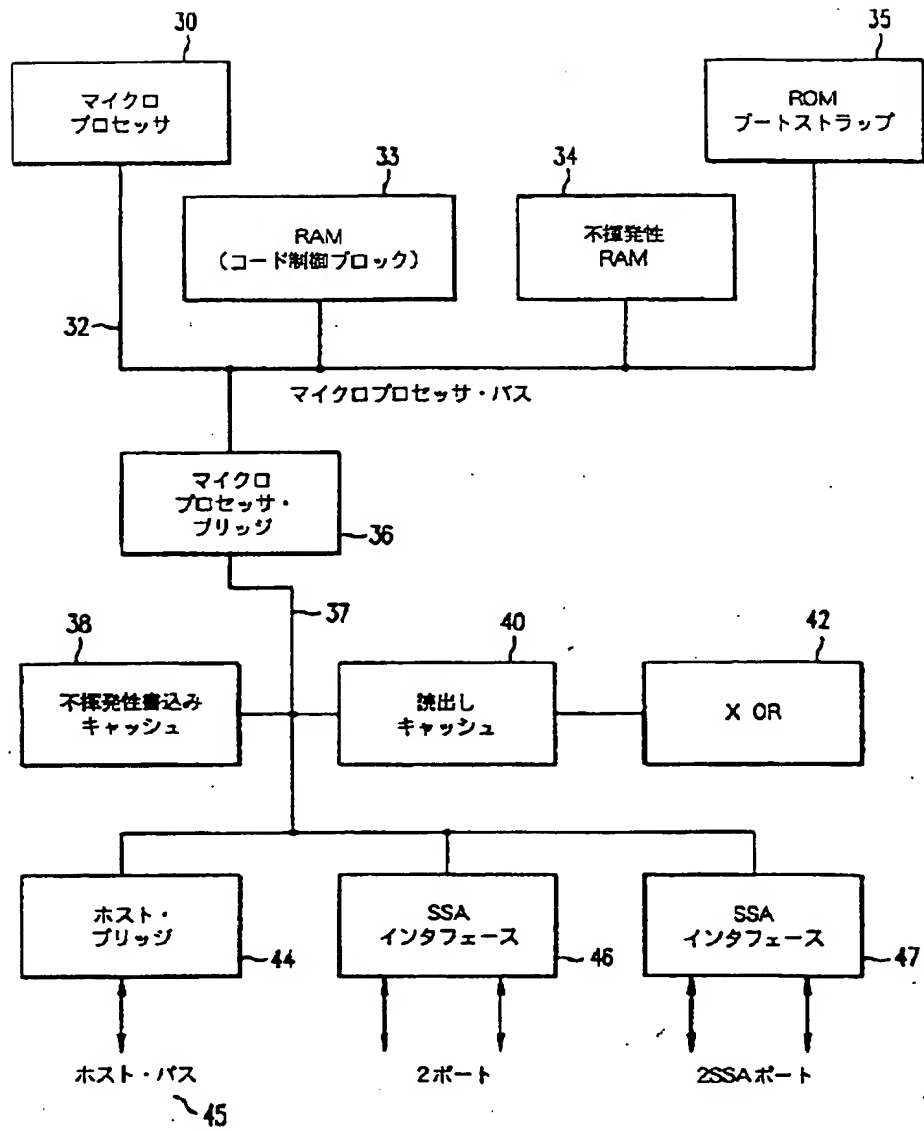
【図3】



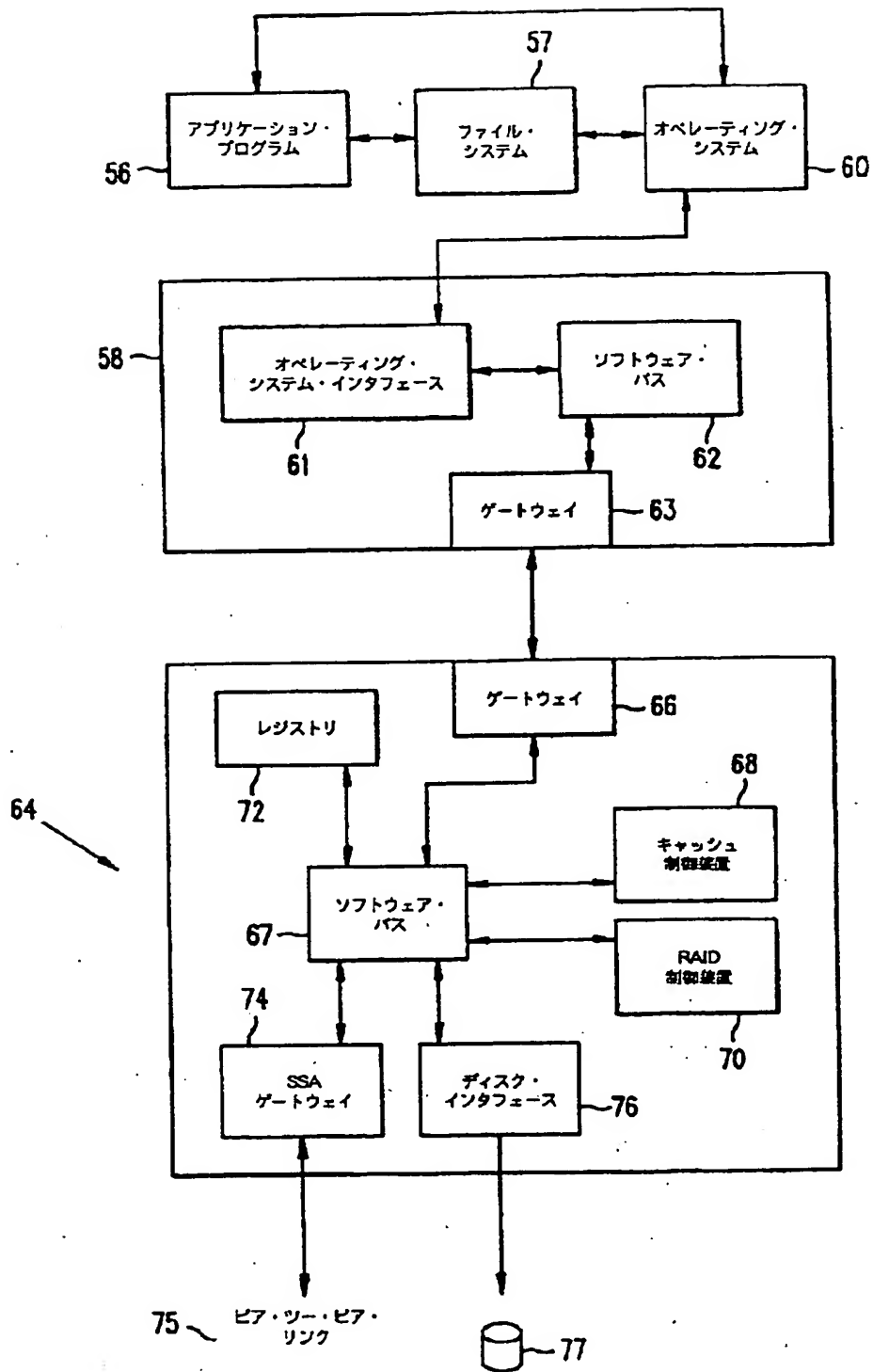
【図4】



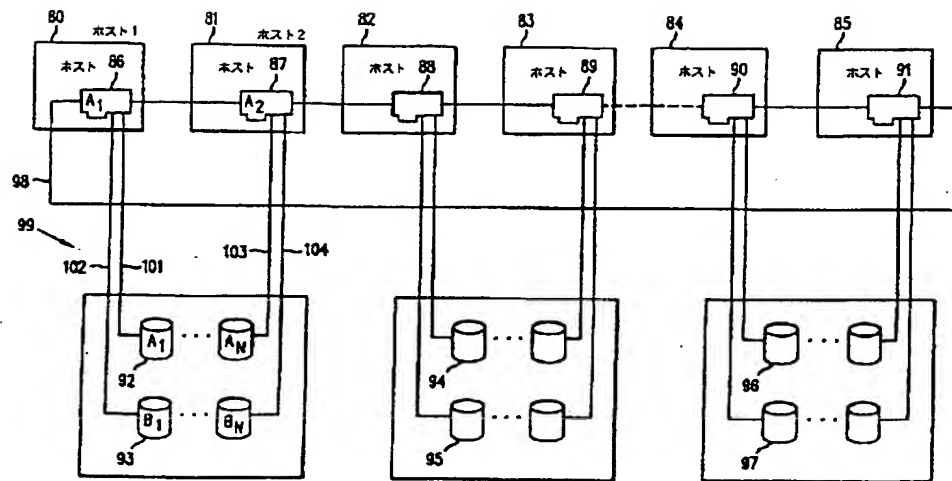
【図5】



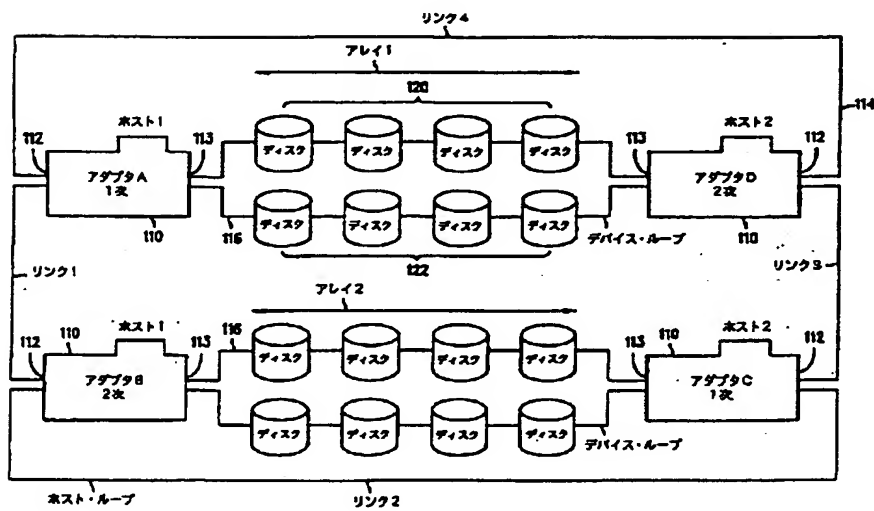
【図6】



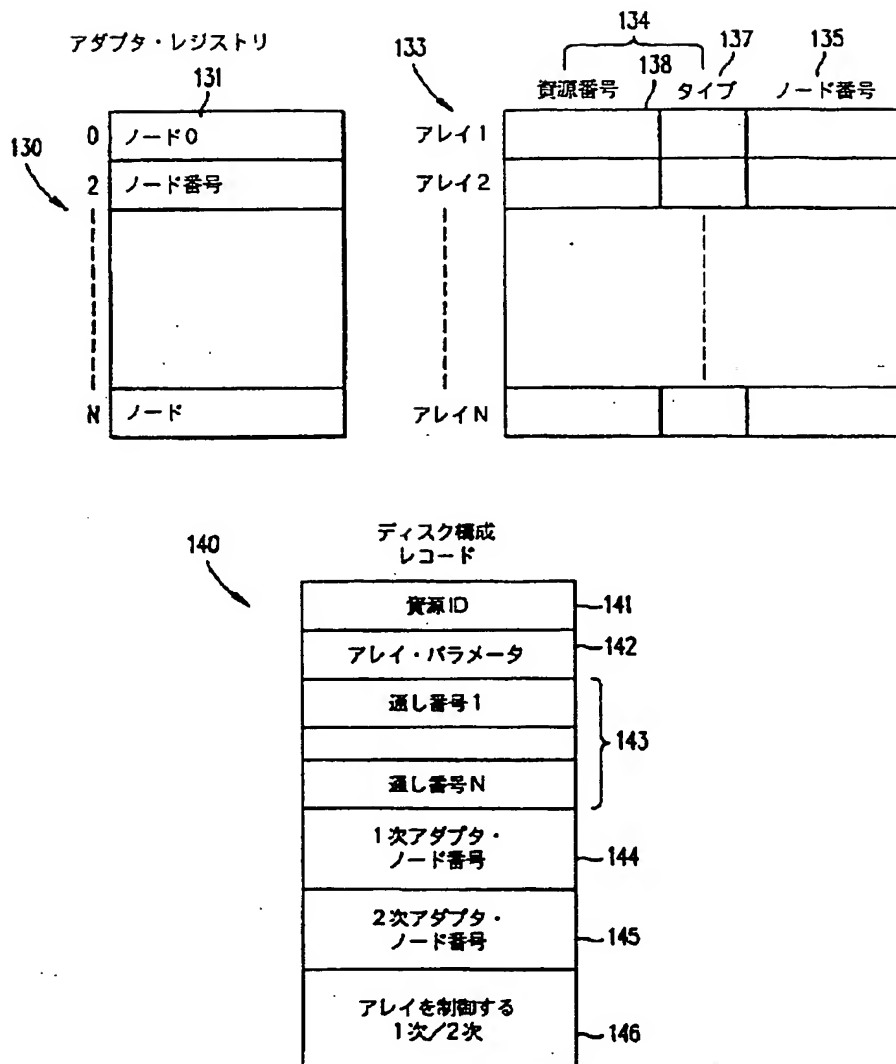
【圖 7】



【图8】



【図9】



フロントページの続き

(72)発明者 ニック・シャイラー
イギリス、エス016 7イー・エフ、サウ
ザンプトン、バセット、グレンカーロン・
ウェイ 14

(72)発明者 アリストアイアー・リースク・サイモン
アメリカ合衆国95123、カリフォルニア州
サン・ホセ、バーチ・リッジ・サークル
309